MARKER DEVELOPMENT FOR QUALITY PROTEIN MAIZE BREEDING AND AN INTERACTION STUDY BETWEEN *Opaque-2* AND *Ask2* GENES

DESARROLLO DE MARCADORES PARA MEJORAMIENTO DE MAÍZ CON PROTEÍNA DE CALIDAD Y ESTUDIO DE INTERACCIÓN ENTRE LOS GENES *Opaco-2* Y *Ask2*

Martha Hernández-Rodríguez[1]*, Raman Babu[2], Debra J. Skinner[2], Natalia Palacios-Rojas[2], Marina C. M. Martins[3], J. Jesús García-Zavala[1], Ricardo Lobato-Ortíz[1], Amalio Santacruz-Varela[1], Fernando Castillo-González[1], Yunbi Xu[2] and Boddupalli M. Prasanna[2]

[1]Colegio de Postgraduados, Campus Montecillo, PREGEP-Genética, Montecillo, Texcoco, Estado de México, México. [2]International Maize and Wheat Improvement Center, Mexico City, México. [3]Max-Planck-Institut für Molekulare Pflanzenphysiologie, Potsdam-Golm, Germany.

*Corresponding author (hernandez.martha@colpos.mx)

SUMMARY

Quality Protein Maize (QPM) kernels contain twice the amounts of lysine and tryptophan compared to normal corn kernels. Although the *opaque-2* (*o2*) mutation is the underlying cause of this beneficial change, other genes such as *Aspartate kinase-2* (*Ask2*) affect the amino acid content in the endosperm to a lesser degree. To date, reports on the interaction between both loci are scarce and there are no high-throughput assays for the identification of the alleles of these genes. The objectives of this research were: 1) to study the interaction between the *o2* and *Ask2* genes with respect to the accumulation of amino acids in the endosperm in an F_2 population, 2) to identify conserved SNPs into the *o2* gene that can be used as markers, 3) to estimate the frequency of an SNP of *Ask2* associated with the accumulation of lysine in the endosperm in CIMMYT germplasm, and 4) to develop high-throughput marker assays for these SNPs. The interaction study showed a preponderant effect of *o2* on the accumulation of 11 amino acids (P ≤ 0.01). *Ask2* appeared only to act with *o2* to enhance marginally lysine, histidine and methionine levels in the double recessive homozygotes. Sequencing of amplicons at the *o2* locus led to the identification of an SNP in exon 1 that discriminated all QPM (C) genotypes from non-QPM (T) genotypes. Validation of this SNP through KASP™ assays indicated that it was 92 % assertive in differentiating the *o2* genotypes. In contrast, the frequency of the *Ask2* SNP in CIMMYT QPM germplasm was low; however, an SSCP marker developed using this SNP detected five variants indicating that other unknown base changes may confer positive lysine-increasing responses. These markers could aid the marker-assisted selection of QPM cultivars.

Index words: *Zea mays*, *Ask2*, marker-assisted selection, *Opaque-2*, SNP.

RESUMEN

Los granos de maíz de alta calidad proteica (QPM) contienen el doble de lisina y triptófano en comparación con los granos de maíz normales. Aunque la mutación *opaco-2* (*o2*) es la causa subyacente de este cambio beneficioso, otros genes como la *aspartato quinasa-2* (*Ask2*) afectan en menor grado el contenido de aminoácidos en el endospermo. Hasta el momento, los informes sobre la interacción entre ambos loci son escasos y no existen ensayos de alto rendimiento para la identificación de los alelos de estos genes. Los objetivos de esta investigación fueron: 1) estudiar la interacción entre los genes *o2* y *Ask2* con respecto a la acumulación de aminoácidos en el endospermo de una población F_2, 2) identificar SNPs conservados en el gen *o2* que puedan ser utilizados como marcadores, 3) estimar la frecuencia de un SNP de *Ask2*, asociado con la acumulación de lisina en el endospermo, en germoplasma del CIMMYT, y 4) desarrollar ensayos de marcadores de alto rendimiento para estos SNPs. El estudio de interacción mostró un efecto preponderante del *o2* sobre la acumulación de 11 aminoácidos (P ≤ 0.01). Al parecer, *Ask2* solamente actúa con *o2* para mejorar marginalmente los niveles de lisina, histidina y metionina en los homocigotos recesivos dobles. La secuenciación de amplicones en el locus *o2* condujo a la identificación de un SNP en el exón 1 que discriminó todos los genotipos QPM (C) de los genotipos no QPM (T). La validación de este SNP a través de los ensayos KASP™ indicó que fue 92 % asertivo en la diferenciación de los genotipos *o2*. En contraste, la frecuencia del SNP de *Ask2* en el germoplasma QPM del CIMMYT fue baja; sin embargo, un marcador SSCP desarrollado utilizando este SNP detectó cinco variantes, lo que indica que otros cambios de base desconocidos pueden conferir respuestas positivas de aumento de lisina. Estos marcadores podrían ayudar a la selección asistida en variedades QPM.

Palabras clave: *Zea mays*, *Ask2*, *Opaco*-2, selección asistida por marcadores, SNP.

INTRODUCTION

There are nine amino acids that human beings and mono-gastric animals cannot synthesize and need to be supplied through the diet (Galili *et al*., 2016). While most proteins of animal origin have suitable proportions of these amino acids, plant proteins are often suboptimal for human needs; for instance, most cereal proteins are poor in lysine and tryptophan, while those of legume origin are lower in methionine and other sulfur-containing amino acids (Mariotti and Gardner, 2019). In the developing world, prolonged consumption and over-reliance on cereal proteins can potentially lead to diseases such as pellagra, a niacin deficiency disease, and kwashiorkor, a form of protein-energy malnutrition that increases susceptibility to tuberculosis and gastroenteritis (Nuss and Tanumihardjo, 2011).

A mature corn kernel contains approximately 10 % protein and several mutants alter its amino acid profile (Zhang *et al.*,

2018). One of them is *opaque-2* (*o2*), a natural spontaneous mutation with soft, opaque grains, discovered in the 1920s (Singleton, 1939). Biochemical effects of *o2* indicate that the homozygous recessive *o2* allele increases lysine (+69 %) and tryptophan in the endosperm compared to normal maize (Mertz *et al*., 1964). This effect doubles the biological value of maize protein, and only with half the amount of *o2* maize (relative to normal maize), the same biologically usable protein can be obtained (Bressani, 1991). In spite of its nutritional superiority, *o2* maize did not become popular among farmers because the *o2* gene carries some negative pleiotropic effects, including grain chalkiness, thicker pericarp, larger germ size, reduced cob weight, lower grain yield and reduced kernel weight and density (Vasal, 2000). Hence, CIMMYT made intensive efforts to improve the phenotype of *o2* kernels by developing hard-endosperm maize types named QPM. Today's QPM is interchangeable with normal maize in both phenotype and yield potential while retaining its nutritional value (Vasal, 2001).

The lysine value in QPM lines varies from 2.5 to 4.0 g per 100 g of protein (Vivek *et al*., 2008). Though *o2* is the major determinant, a few minor modifier loci may likely work in unison with *o2* to moderate the effect and alter the essential amino acid content in the homozygous (*o2o2*) recessive background (Vasal, 2001). In a QTL mapping study by Wang and Larkins (2001), free amino acid content in *o2* maize lines was analyzed and shown to vary up to 12-fold in Oh545*o2* and 3-10 times in Oh51A*o2* and W64A*o2* from non-*o2* wild type counterpart. One of four loci identified was coincident with the *Aspartate kinase 2* (*Ask2*) gene (Wang *et al*., 2001). This locus, located on chromosome 2, has a polymorphism that causes a single amino acid substitution in the C-terminal region of the ASK2 enzyme of Oh545*o2* (Wang *et al*., 2007). Another study using a RIL population from the cross between B73*o2* and CML161, identified QTLs for traits related to amino acid content on chromosomes 7 and 8, explaining up to 39 % of the observed phenotypic variation (Gutiérrez-Rojas *et al*., 2010). These studies clearly established the existence of additional modifier loci that could favorably increase accumulation of amino acid levels in QPM germplasm. Of the several loci reported, the role of *Ask2* is prominent and the responsible polymorphism has been identified and characterized (Wang *et al*., 2007).

The cloning and characterization of the *O2* locus, located on chromosome 7 (Hartings *et al*., 1989; Schmidt *et al*., 1990), followed by detection of three SSR markers *viz*. phi057, phi112 and umc1066 (Chin *et al*., 1996) led to effective differentiation of the *O2* and *o2* alleles in elite QPM and non-QPM germplasm (Babu *et al*., 2005). These *o2*-specific SSR markers are the foundation for marker-assisted selection (MAS) in QPM breeding (Babu *et al*.,

2005; Gupta *et al*., 2013; Pukalenthy *et al*., 2020); however, they are not enough to achieve the full effectiveness of molecular breeding for QPM genotypes. Marker umc1066 is easily visualized on agarose gels but it is commonly not polymorphic, at least in CIMMYT breeding populations; phi057 usually requires the use of polyacrylamide gels, while phi112, which is based on a deletion in the promoter region, is a dominant marker, and hence, cannot be used in the identification of heterozygotes in F_2/BC generations (Babu and Prassana, 2014). These shortcomings, coupled with low throughput and time consuming of gel-based SSR analysis, result in a major hindrance to routine and large-scale adoption of marker technology in the breeding pipelines (Xu and Crouch, 2008). This requires the identification of SNP markers, which have propelled plant molecular breeding because of their ease of automation, high throughput, low cost and high read accuracy (Rasheed *et al*., 2017). Though the functional SNP conferring favorable response in *Ask2* has been reported (Wang *et al*., 2007), high throughput assays are not yet available and the interaction between *o2* and *Ask2* loci, involving breeding germplasm, has not been documented. The objectives of this study were: 1) to evaluate the interaction between *o2* and *Ask2* in an F_2 population derived from tropical maize germplasm and validate synergistic effect, if any, for enhanced amino acid accumulation in the endosperm, 2) to identify widely conserved SNPs within the *o2* gene that can effectively discriminate the alleles in a range of elite genetic backgrounds, 3) to estimate the frequency of favorable *Ask2* allele in CIMMYT germplasm and, 4) to develop medium to high throughput marker assays to facilitate the implementation of a comprehensive MAS scheme for rapid development of QPM germplasm.

MATERIALS AND METHODS

Interaction analysis

Nine genotypic classes were determined after genotyping 400 seeds of an F_2 population resulting from a cross between CML172 (QPM) and CML323 (non-QPM). Genotyping was carried out with molecular markers for *o2* and *Ask2* using DNA from endosperm. DNA extraction followed a protocol based on CTAB-lauroylsarcosine described by Gao *et al*. (2008). To genotype *o2*, SSR marker phi057 was amplified by PCR reactions of 15 µL containing 1X Taq buffer, 2.5 mM $MgCl_2$, 150 µM each dNTP, 0.3 µM of each primer (forward 5´-CTCATCAGTGCCGTCGTCCAT-3´ and reverse 5´-CAGTCGCAAGAAACCGTTGCC-3´), 1 U homemade Taq polymerase and 20 ng of DNA. The PCR program consisted of an initial denaturation of 1 min at 94 °C followed by 30 cycles of 30 s at 94 °C, 1 min at 58 °C, 1 min at 72 °C and a final extension of 5 min at 72 °C using a MJ Tetrad thermalcycler (MJ Research, St. Bruno, Quebec,

Canada). Three microliters of each PCR product was migrated in 6 % acrylamide:bis-acrylamide (29:1) gels at 250 V for 2 h in 1X TBE (90 mM Tris-borate, 2 mM EDTA pH 8.0) in a MGV-116-33 system (CBS Scientific®, San Diego, California, USA). The reference of molecular weight was φX174/*Hae*III. After electrophoresis, the gels were silver-stained according to CIMMYT (2005).

To genotype *Ask2*, primers flanking the functional SNP of *Ask2* (Wang *et al.*, 2007) were designed to amplify a product of 151 pb. PCR mix included 1X Taq buffer, 2.0 mM $MgCl_2$, 166 µM each dNTP, 0.1 µM of each primer (forward 5´-CTTGGTGGTCCATGACAGTG-3´ and reverse 5´-CTAATGGCTGTGGATTGTGC-3'), 1 U homemade Taq polymerase and 50 ng of DNA in a reaction volume of 15 µL. The PCR program consisted of one cycle of 94 °C for 2 min, 30 cycles at 94 °C for 1 min, 58°C for 40 s, 72°C for 1 min, and a final extension at 72 °C for 5 min performed in a MJ Tetrad thermalcycler (MJ Research, St. Bruno, Quebec, Canada). PCR products were resolved as single-stranded conformation polymorphisms (SSCP) in 16 % polyacrylamide gels applying a modified Laemmli´s protocol that consisted of preparing each layer of the gel (stacking and resolving) without SDS (Laemmli, 1970). The PCR products were diluted with 5 µL of loading buffer (10 mM NaOH, 95 % formamide, 0.05 % bromophenol blue, 0.05 % xylene cyanol), denatured during 10 min at 94 °C and quick quenching on ice. Five µL were run for 15-16 h at 150 V and 8 °C in a MGV-116-33 system (CBS Scientific®, San Diego, California, USA) using 1X TBE as running buffer. The gels were stained with 1X SYBR® Gold (Life Technologies, Carlsbad City, California, USA) for 20 min and documented by scanning in a Typhoon 8600 imager (Amersham Biosciences Ltd., Chicago, Illinois, USA).

After genotyping, each marker was scored by hand and seeds with the same genotype were bulked to form a genotypic class. Only 20 seeds per class were used to quantify 11 amino acids by high-performance liquid chromatography (HPLC). Ethanolic extracts by HPLC (consisting of a pump P680, Dionex, autosampler 465 and fluorimetric detector SFM 25, Kontron, Eching, Germany) after derivation of the primary amino group with *o*-phthaldialdehyde (OPA) were carried out as described by Jarrett *et al.* (1986). Aliquots (35 µL) of ethanolic extracts were combined with 35 µL of OPA reagent [25 mg OPA dissolved in 500 µL methanol, the addition of 4.5 mL 0.8 M borate buffer (pH 10.4) and 50 µL of mercaptopropionic acid], mixed and kept for 2.5 min at 4 °C to allow complete derivation. The extract was injected into a pre-column (Hypersil 10 µm, Phenomenex, Torrance City, California, USA) before separation on a C18 column (Hypersil ODS, 150 × 4.6 mm, 3 µm particle size, Thermo Fisher Scientific, Bonn, Germany) maintained at constant 18 °C using a

thermostatic column holder (560-CIL, ERC, Riemerling, Germany) at a flow rate of 0.8 mL min^{-1}. Eluent A was 8 mM Na_2HPO_4 (pH 6.8) containing 0.4 % (v/v) tetrahydrofuran. Eluent B was a mixture of 250 mL 40 mM Na_2HPO_4 (pH 6.8), 150 mL methanol and 105 mL acetonitrile. Two different gradients determined major and minor amino acids. For major amino acid columns were equilibrated for 10 min with 10 % A and separation conditions were as follows: 0-5 min, isocratic elution with 0 % B; 5-15 min, linear gradient from 0 % to 15 % B; 15-15.5 min, linear gradient from 15 to 100 % B; 15.5-21 min 100 % B. For minor amino acids, columns equilibrated for 10 min with 100 % A and separation conditions were as follows: 0-5 min, isocratic elution with 0 % B; 5-16 min, linear gradient from 0 to 15 % B; 16-26 min, linear gradient from 15 to 50 % B; 26-42 min, linear gradient from 50 to 60 % B; 42-46 min from 60 to 100 % B; 46-52 min 100 % B. Amino acids were identified and quantified by comparison with original standards (Sigma-Aldrich) using Chromeleon software (version 6.8, Dionex). Individual analyses of variance per amino acid and principal component analysis were applied using PROC GLM and PRINCOMP of SAS, respectively (SAS Institute, 2002).

Polymorphism identification

Fifteen elite inbred lines were chosen to search for polymorphisms within the *o2* gene. Seven inbred lines were QPM, of diverse agro-ecological zones, with excellent kernel modifying capacity and of extensive use in the QPM conversion programs. The other eight lines were non-QPM, with wide adaptability and superior performance in Asia, Latin America and East African highlands (Table 1). Leaves of four seedlings per line were used separately to extract genomic DNA following a protocol based on CTAB for small-scale isolation of high quality DNA (CIMMYT, 2005). Two sets of primers were designed to amplify two genomic fragments of *o2* (477 and 530 pb) and to identify regions with high degree of conservation. The reference sequence X15544, which belongs to the zein regulatory gene *o2* of *Zea mays* (http://www.ncbi.nlm.nih.gov/), plus six cDNA sequences of *o2* gene (AJ41297, AJ491298, AJ491299, AJ491300, M29411 and X16618) were used as templates for the design of primers O2F1: 5´-CTAGTGTTTGCTTCTCCCTTCC-3´, O2R1: 5´-TCCTCAGTATGGCATTGTACTCC-3´ and O2F3: 5´-TGTTGTGACCTCAGATCAACG-3´, O2R3: 5´-TTCCAGTTCTTTCAGGTGAGC-3. The PCR reactions were optimized to produce single bands without smear by applying a lower number of cycles and a large volume of reaction (30 µL). In total, two PCR fragments per seedling were amplified and purified with Qiagen spin columns (Qiagen®, Hilden, Germany), and cloned into pGEM®-T Easy Vector (Promega, USA) according to manufacturer recommendations. Recombinant plasmids were selected

Table 1. Maize inbred lines used to identify polymorphisms in the *o2* locus.

Grain color and adaptation zone	QPM line	Non-QPM line
White, tropical lowland	CML159	CML343
	CML491	CML495
	CML502	CL RCW22
	CML492	
Yellow, tropical lowland	CML161	CML348
	CML165	
White, sub-tropical	CML176	
White, highland		CML244
		CML349
White, Ethiopian highland		F7287
		A7018

after transforming DH5α competent cells of *E. coli* by heat-shock according to CIMMYT (2005). Plasmid DNA extraction was with Wizard® *Plus* SV Minipreps DNA Purification System (Promega, USA) and verification of inserts was by digestion with *Eco*R1. Two plasmids of each genotype were sequenced at CINVESTAV facilities (Guanajuato, México), aligned using Clustal X version 1.8 (Thompson *et al.*, 1997) and refined by hand.

Markers assays validation

A panel of 88 inbred lines from the Maize Program or CIMMYT Genbank was used to validate polymorphisms using genomic DNA as template. DNA extraction protocol was the same as that used for polymorphism identification. Two SNPs were assayed at LGC Biosearch Technologies (https://www.biosearchtech.com/) using the KASP™ chemistry. One of them was the *o2*-specific SNP, which differentiated QPM and normal lines and the other was the *Ask2*-specific SNP amplified to determine genotypic classes. This last one was also screened as SSCP marker to explore the natural variation of *Ask2* and to estimate the frequency of the favorable allele in CIMMYT germplasm.

RESULTS

Interaction analysis between *o2* and *Ask2*

The interaction between *o2* and *Ask2* was studied in a tropical germplasm background by developing a segregating population from a cross between CML172, which is a QPM line possessing favorable alleles for both *o2* and *Ask2* loci (based on '2' type SSCP conformer), and CML323, a non-QPM line. The interaction analysis was based on the concentration of 11 amino acids of nine genotypic classes (Table 2). Among these classes, the double recessive homozygous class (G1) contained

significantly higher amounts of lysine, tryptophan and aspartic acid relative to the double dominant homozygous class (G5). In this regard, G1 had 2.3-fold more lysine, 1.3-fold more tryptophan and 5.3-fold more aspartic acid than G5. Lysine levels of G1 were also 60 % higher relative to G2, the recessive dominant class. A fast analysis using biplot analysis indicated that the first two components explained 96.24 % of the variance (Figure 1). PC1 explained 89.27 % of variation while PC2 explained 6.93 %. All vectors had a positive direction toward G1, G2 and G3 where *o2* was recessive. No vector pointed toward genotypes with the dominant allele of *o2* (G4 to G9). On the contrary, *Ask2* locus on its own was not found to have any significant effect on any of the amino acids analyzed, but appeared to interact synergistically with *o2* and it significantly enhanced lysine, histidine and methionine levels in the double recessive homozygous class (G1).

Polymorphism identification in the *o2* locus

Sequencing the amplicons of *o2* allowed to locate an important discriminatory SNP in the region of exon 1 (Figure 2). The sequencing included part of the 5' flanking region, four exons (1, 2, 3 and 4) and two introns (2 and 3). In total, a length of 1 kb of the genomic sequence was analyzed, 65 polymorphic sites were identified, and one polymorphism each 30 bp was averaged. In general, all QPM lines had the same sequence, indicating a possible common donor of *o2* in CIMMYT lines.

Of the several polymorphisms that differentiated the QPM lines from non-QPM, the following four merit attention: 1) two SNP in intron 2 (positions 10,800,027 and 10,800,032 bp concerning to *o2* reference sequence of *Zea mays*, ID 542375 of B73 RefSeq_v3) which defined three alleles with the haplotypes GT/AG/GT (donor/recipient/recipient), 2) two consecutive SNP in intron 2 (positions 10,800,038 and 10,800,039) which defined the alleles AT/TC/AC (donor/recipient/recipient), 3) two InDels in intron 2 (positions 10,799,978 to 10,799,982 and 10,799,988 to 10,799,990), which corresponded to 5-base and 3-base deletions in all QPM donor lines as well as for two non-QPM lines, and 4) an SNP of transition type in exon 1 (position 10,799,047 according to sequence ID 542375 of B73 RefSeq_v3, and position 1790 according to accession X15544 of the GenBank). The SNP corresponds to C in QPM lines and to T in non-QPM lines. This SNP was named 2-exon1-SNP, it causes an amino acid substitution (valine to alanine) and was selected for the subsequent high throughput assays. Sequences with this SNP were submitted to the GenBank of NCBI under accessions KY807178 to KY807200.

Table 2. Amino acids concentration (nmol g⁻¹ FW) of kernels of the genotypic classes of *o2* and *Ask2* in CML172 × CML323.

Genotypic class	o2	Ask2	Asp	Glu	Asn	Thr	His	Met	Trp	Phe	Ile	Leu	Lys
CML172	1	1	4969.74	2759.53	4631.05	286.31	61.10	27.98	53.14	210.12	159.79	268.78	719.06
CML323	2	2	268.04	329.00	656.81	64.00	41.36	39.11	45.95	36.17	41.80	55.04	81.75
G1	1	1	3383.40	1203.85	1469.11	297.81	230.78	100.67	105.13	155.70	199.92	224.46	431.92
G2	1	2	3095.29	1294.23	2312.75	349.47	160.47	47.47	112.76	156.63	192.91	249.35	270.02
G3	1	H	1979.50	1108.96	1189.26	306.18	120.90	60.67	66.57	112.90	169.72	250.83	364.71
G4	2	1	409.54	393.07	351.05	179.56	68.63	31.10	37.67	58.01	81.12	171.76	94.02
G5	2	2	541.97	557.88	589.25	222.23	98.03	36.49	46.35	67.43	94.24	177.15	132.78
G6	2	H	460.78	366.26	374.36	180.02	104.14	38.35	49.94	48.73	62.50	129.35	124.24
G7	H	1	298.44	247.67	309.08	149.93	73.33	32.13	37.31	46.06	46.36	103.12	102.82
G8	H	2	471.34	342.96	385.82	170.18	95.32	39.26	38.24	50.65	72.13	128.60	126.01
G9	H	H	315.77	300.76	300.61	159.87	88.22	31.29	39.22	43.81	57.83	114.87	96.99
o2 classes													
G1,G2,G3	1	-	2819.40[a]	1202.35 [a]	1657.04 [a]	317.82 [a]	170.72	69.61	94.82 [a]	141.74 [a]	187.52 [a]	241.55 [a]	355.55 [a]
G4,G5,G6	2	-	470.76[b]	439.07[b]	438.22[b]	193.94[b]	90.27	35.31	44.66[b]	58.05[b]	79.29[b]	159.42[b]	117.01[b]
G7,G8,G9	H	-	361.85[b]	297.13[b]	331.83[b]	159.99[c]	85.62	34.23	38.25[b]	46.84[b]	58.77[b]	115.53[b]	108.61[b]
Ask-2 classes													
G1,G4,G7	-	1	1363.79	614.86	709.74	209.10 [a]	124.25	54.64	60.04	86.59	109.13	166.45	209.59
G2,G5,G8	-	2	1369.53	731.69	1095.94	247.30[b]	117.94	41.08	65.78	91.57	119.76	185.04	176.27
G3,G6,G9	-	H	918.69	591.99	621.41	215.36[b]	104.42	43.44	51.91	68.48	96.68	165.02	195.31

Genotype nomenclature: 1: double homozygous recessive, 2: double homozygous dominant, H: heterozygous. Asp: aspartic acid, Glu: glutamic acid, Asn: asparagine, Thr: threonine, His: histidine, Met: methionine, Trp: tryptophan, Phe: phenylalanine, Ile: Isoleucine, Leu: leucine, Lys: lysine. Means with different letters within columns are statistically different (Tukey, P ≤ 0.05).
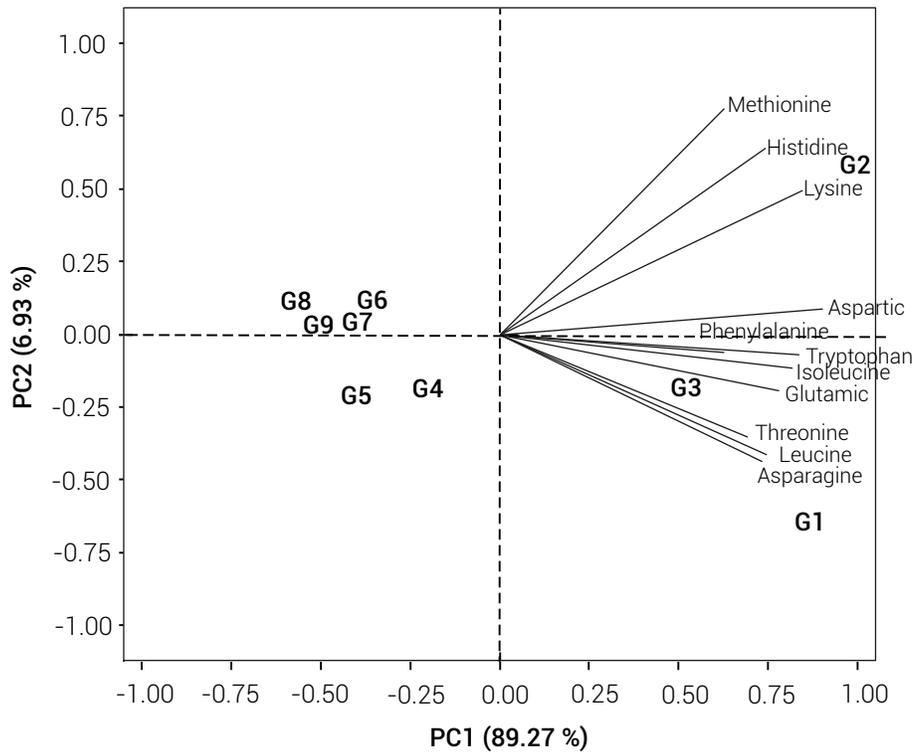
Figure 1. Biplot of principal components of 11 free amino acids content in nine genotypic classes (G) from an F$_2$ population between CML172 (QPM) and CML323 (non-QPM). G1: *o2/o2|ask2/ask2*, G2: *o2/o2|Ask2/Ask2*, G3: *o2/o2/Ask2/ask2*, G4: *O2/O2/ask2/ask2*, G5: *O2/O2/Ask2/Ask2*, G6: *O2/O2/Ask2/ask2*, G7: *O2/o2/ask2/ask2*, G8: *O2/o2/Ask2/Ask2*, G9: *O2/o2/Ask2/ask2*.
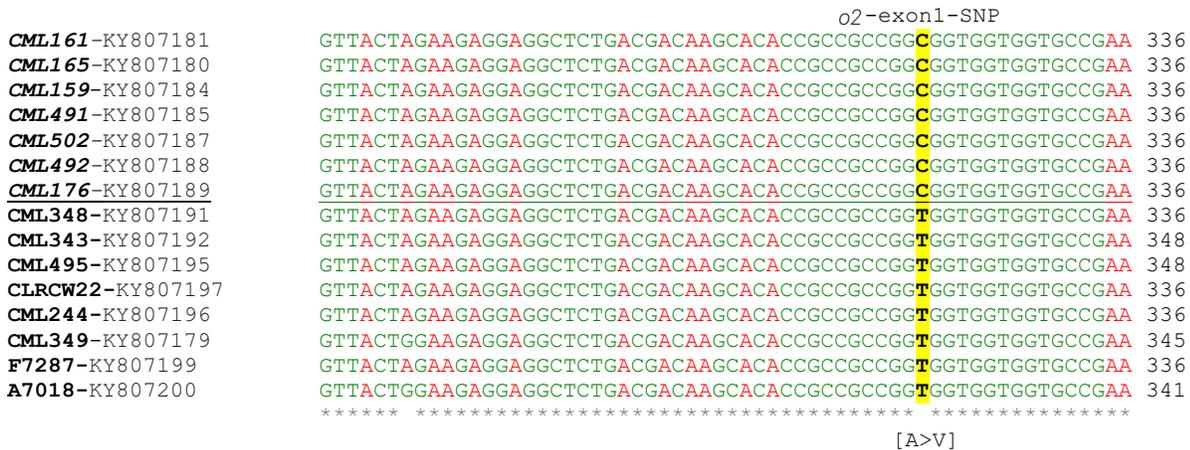

Figure 2. The *o2*-exon1-SNP and its context sequence. The SNP is highlighted in yellow. Allele C discriminates to QPM lines (in italics). Allele T discriminates non-QPM lines (no italics). Name of each maize line plus its GenBank ID is to the left of the nucleotide sequence.

Validation of high throughput marker assays

Two high throughput KASP™ assays non gel-based were developed to facilitate large-scale MAS in pedigree/back-cross breeding programs involving QPM × non-QPM lines. One of these assays, designed to screen the *o2*-exon1-SNP, was firstly tested in 16 lines. The analysis of allele calls clearly grouped all genotypes into two clusters −one corresponding to allele C of QPM and another corresponding to allele T of non-QPM lines. Thus, this analysis was extended to a larger set of 88 lines in which 60 lines were QPM and the remaining non-QPM. Allele calls for *o2*-exon1-SNP differentiated two clusters again (Figure 3A), the cluster in red color corresponding to lines with the allele C and the cluster in blue color to lines with the allele T. The cluster with allele C included 53 QPM lines while the cluster with allele T comprised 35 lines, of which 28 were normal and seven were QPM. CML145, CML167, CML168, CML177, CML180, CML184 and CML188 were the QPM lines that possessed the 'normal' allele (T), contrary to expectation. Therefore, prediction value of *o2*-exon1-SNP was 92 % for the QPM trait.

The remaining assay was designed to test the functional polymorphism A ('lysine enhancing') of the C-terminal region of the ASK2 protein of the Oh545o2 maize line (Wang *et al.*, 2007), which identified several lines possessing the favorable allele (Figure 3B). In this case, allele calls distinguished one cluster (in red) of seven inbred lines with the favorable allele A, of which five were QPM (CML148, CML161, CML162, CML179 and CML188) and two normal

(CML328 and CML 370). The other cluster with unfavorable allele T (in blue) comprised 81 lines of which 55 were QPM and 26 non-QPM, indicating a low frequency of this allele in CIMMYT germplasm.

Natural variation of *Ask2*

Given that the favorable allele A at the *Ask2* locus was only detected in five out of 60 QPM lines, to determine other possible functional variants of *Ask2*, a gel-based SSCP assay was developed and tested in 82 maize lines. Five distinct SSCP band patterns were unambiguously identified (Figure 4). SSCP variants were as follows: variant 1 in 60 lines (46 QPM and 14 normal), variant 2 in 11 lines (10 QPM and 1 normal), variant 3 in 5 lines (all QPM), variant 4 in 5 lines (all normal) and variant 5 in 1 line (QPM), suggesting that other polymorphisms could be involved in the high lysine level in the endosperm.

DISCUSSION

The interaction between *o2* and *Ask2* in tropical germplasm background was studied in nine different genotypic classes resulting from an F_2 segregating population. The single locus tests established the overriding effect of *o2* on most of the amino acids, irrespective of the genotypic constitution at *Ask2* locus (Figure 1). On the contrary, *Ask2* locus on its own was not found to have any significant effect on any of the amino acids analyzed but did interact synergistically with *o2* and significantly enhanced lysine, histidine and methionine levels in the
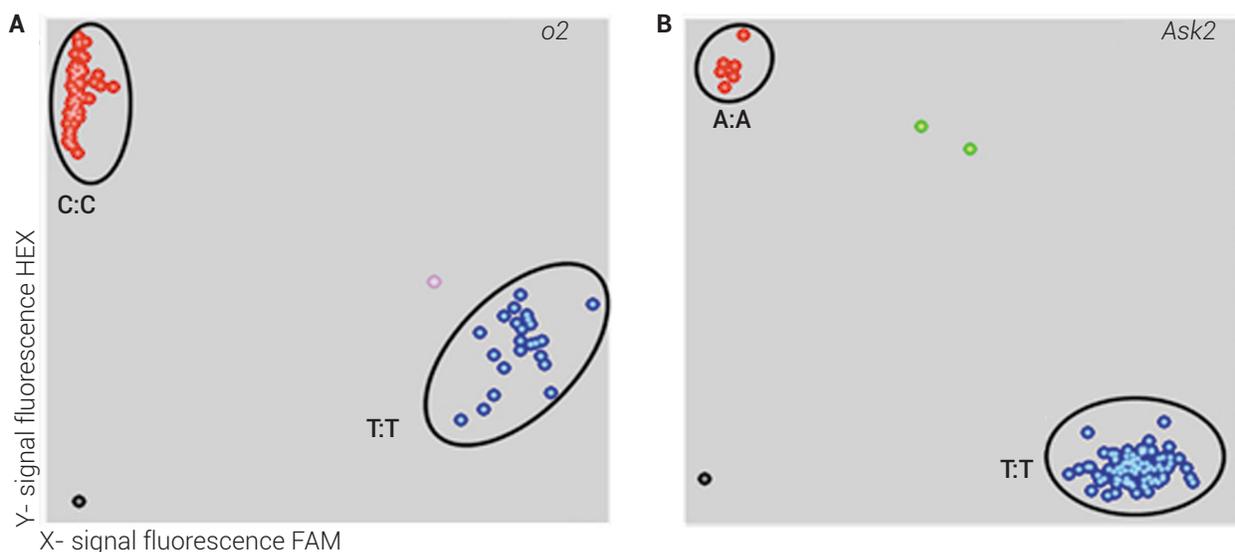

Figure 3. Genotyping cluster plot of KASP™ assays for *o2* (A, left) and *Ask2* (B, right) SNP. Each data point represents an individual inbreed line. Similar fluorescence values mean the same genotype. Red dots: homozygous for allele labeled with HEX, blue dots: homozygous for allele labeled with FAM, green dots: heterozygous, black dots: no template control.
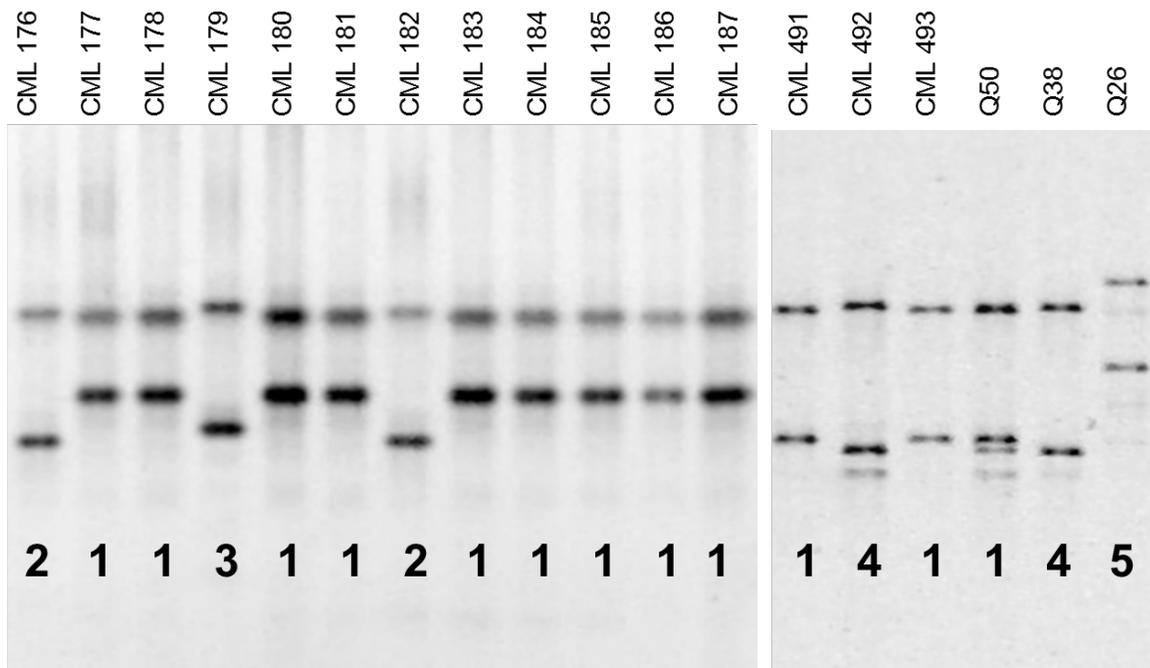
Figure 4. The five SSCP variants of *Ask2* detected in QPM germplasm. Mobility differences correspond to base changes in the amplified sequence (151 pb, from 3' end) that cause single-stranded DNA to migrate differently. Bands were resolved on 16 % polyacrylamide gel.

double recessive homozygous class (G1).

The elevated levels of lysine in *o2* endosperm are due to enhanced synthesis of lysine-containing proteins (Zhang *et al*., 2018) as well as high free lysine content (Wang and Larkins, 2001). Lysine is synthesized through a branch of aspartate pathway and is regulated by feedback inhibition loops, in which the accumulating lysine suppress the activity of two key enzymes, aspartate kinase (AK) and dihydrodipicolinate synthase (DHDPS) (Galili, 1995; Galili *et al*., 2016). The native mono-functional AK enzyme is a tetramer composed of two α- and two β-subunits, encoded by *Ask1* and *Ask2* genes, respectively (Dotson *et al*., 1989; 1990). Mutant alleles in either *Ask1* or *Ask2* loci result in AK enzyme that is less sensitive to feedback inhibition and lead to overproduction of lysine, threonine, methionine and isoleucine (Dotson *et al*., 1990). In maize, *Ask2* explains a significant portion of variance for the free lysine levels in the *o2* endosperm (Wang and Larkins, 2001), and thus, it represents an important target for enhancing free lysine levels in seeds. In this study, the mutation in the *Ask2* locus that resulted in a favorable SSCP allele in CML172 is either insensitive or less sensitive to lysine inhibition and, thereby, contributed to increased accumulation of lysine and other aspartate pathway amino acids, similar to the mutant *Ask2* allele of Oh545*o2*; however, it was apparent that the *Ask2* locus on its own did not have any significant impact on the lysine or other amino acid levels in this investigation. Based on studies involving double mutants of *Ask1* and *o2*,

Brennecke *et al*. (1996) inferred that *Ask1* is being directly or indirectly regulated by *O2* because AK in these double mutants is less sensitive to lysine feedback inhibition than that of *Ask1* mutation alone. Though there is no direct evidence for *Ask2* being regulated by *O2*, such a possibility may partially explain the insignificant effect of *Ask2* locus independent of *o2*.

Though the three *o2*-specific SSR markers offer an excellent opportunity for MAS, their use has a limited potential in plant breeding. In this investigation, an important discriminatory SNP in the exon 1 region of *o2* was identified and an easy to use, high throughput assay, non-gel-based to facilitate large-scale MAS was developed. The lack of a perfect correlation between the underlying allelic state at this SNP locus and the QPM status indicates the non-causal nature of this variation, which however has a high predictive value for the QPM trait. Assays SSCP and KASP were designed for *Ask2* which are able to screen a large collection of QPM and normal lines in the CIMMYT germplasm and to identify several lines possessing the favorable allele. Interestingly, the frequency of the unfavorable allele at the *Ask2* locus was significantly higher in the QPM germplasm compared to normal endosperm, which could be due to the high selection pressure exerted by breeders during the process of QPM germplasm development through rigorous biochemical tests for high lysine and tryptophan. The higher free amino acid content in the *o2* endosperm could either be due to increased lysine

synthesis or loss of lysine ketoglutarate reductase (LKR) activity, which degrades lysine as endosperm matures (Arruda *et al.*, 2000). Apart from *Ask2*, several genes have been postulated to play key roles in the enhanced lysine synthesis such as aspartate kinase-homoserine dehydrogenase-2 and dihydro dipicolinate synthase (Galili *et al.*, 2016; Wang and Larkins, 2001). Some of such loci may likely be at play in those QPM lines that are rich in lysine but do not have the favorable *Ask2* allele. Identifying and characterizing such loci in the high lysine QPM lines may provide further incremental gains on the enhanced amino acid levels conferred by *o2* and *Ask2*.

The gel-based SNP detection assay developed in this investigation for the *Ask2* locus combines reasonable throughput of modern gel electrophoresis platforms with robust discriminative power of the SSCP technique (Orita *et al.*, 1989). The SSCP-SNP assay is efficient and sensitive enough for most of the PCR products in the size range of 100-750 bp, cost-competitive and requires no sophisticated instrumentation (Sunnucks *et al.*, 2000). Notably, some of the lines (*e.g.* CML172) that were identified as possessing favorable allele at the *Ask2* locus ('2' type), as per SSCP conformers, had indeed the unfavorable allele (T) according to KASP genotype calls. This discrepancy between SSCP and KASP genotypes in the *Ask2* locus was possibly due to other unknown base changes in the *Ask2* amplified region. Based on the results of these genetic analyses in the segregating F_2 progenies, it was demonstrated that unknown base changes in the exon region of *Ask2* may also confer positive responses (lysine enhancing) similar to the previously established functional SNP. It is indeed, possible to obtain higher genetic gains for enhanced lysine and tryptophan levels if the QPM lines identified in this investigation, that harbor favorable alleles (either A as per KASP assay or '2' type SSCP conformer) at the *Ask2* locus, are preferred as QPM donors in the new line conversions or pedigree starts and tracked through MAS in the subsequent multiple filial generations.

With the development of an array of low-cost and high-throughput genotyping platforms (Semagn *et al.*, 2014), MAS has emerged as an attractive proposition in most commercial and large-scale public-sector maize breeding programs and quickly becoming a compelling option especially ever since doubled haploids are integrated as a mainstream breeding tool in the maize improvement pipeline (Chaikam *et al.*, 2019; Maqbool *et al.*, 2020). Four possible ways are envisioned in which the high throughput SNP markers developed in this investigation for *o2* and *Ask2* could be deployed in the QPM breeding programs: 1) as an inexpensive allele mining tool to identify putative donor lines in large and diverse germplasm collections, including non-adapted genotypes and wild relatives that

have never been subjected to biochemical tests, 2) as an informed choice of parental lines based on the favorable allele composition and, consequently, as a predictive tool to cut down the number of crosses in QPM breeding program, 3) to fix the favorable alleles at *o2* and *Ask2* loci in homozygous condition at an early stage of recombination such as $F_2/BC_1/BC_1F_2$ in pedigree/backcross breeding schemes, and 4) as fixed effect covariates in a genomic selection index to efficiently predict the likely high-lysine/tryptophan entries among the biochemically untested germplasm such as conventionally bred material or DH lines that are generated routinely in large numbers in the breeding program.

CONCLUSIONS

Results pointed to an independent effect of *o2* and *Ask2* rather than an interaction effect between both loci concerning to amino acids content in the maize endosperm. The *o2* variant alone provided a highly significant effect for high accumulation of amino acids in the double recessive genotypes and this accumulation could be positively influenced and modified for lysine, histidine and methionine by *Ask2*. Evidence is provided from high-throughput assays that could enhance the efficacy of marker selection for QPM germplasm development. One of them, the SNP of exon 1 of the *o2* locus had a highly predictive diagnostic value for the QPM trait, while the other one, the SSCP marker led to the identification of genetic variants in QPM germplasm that opens the opportunity to find other polymorphisms associated with the high accumulation of lysine.

ACKNOWLEDGMENTS

BIBLIOGRAPHY

Arruda P., E. L. Kemper, F. Papes and A. Leite (2000) Regulation of lysine catabolism in higher plants. *Trends in Plant Science* 5:324-330, https://doi.org/10.1016/S1360-1385(00)01688-5

Babu R., S. K. Nair, A. Kumar, S. Venkatesh, J. C. Sekhar, N. N. Singh, ... and H. S. Gupta (2005) Two-generation marker-aided backcrossing for rapid conversion of normal maize lines to quality protein maize (QPM). *Theoretical and Applied Genetics* 111:888-897, https://doi.org/10.1007/s00122-005-0011-6

Babu R. and B. M. Prassana (2014) Molecular breeding for quality protein maize (QPM). *In*: Genomics of Plant Genetic Resources. Vol. 2. Crop Productivity, Food Security and Nutritional Quality. R. Tuberosa, A. Graner and E. A. Frison (eds.). Springer. Dordrecht, The Netherlands. pp:489-505, https://doi.org/10.1007/978-94-

007-7575-6_21

Brennecke K., A. J. Souza Neto, J. Lugli, P. J. Lea and R. A. Azevedo (1996) Aspartate kinase in the maize mutants *ASK1-LT19* and *OPAQUE-2*. *Phytochemistry* 41:707-712, https://doi.org/10.1016/0031-9422(95)00634-6

Bressani R. (1991) Protein quality of high-lysine maize for humans. *Cereal Foods World* 36:806-811.

Chaikam V., W. Molenaar, A. E. Melchinger and P. M. Boddupalli (2019) Doubled haploid technology for line development in maize: technical advances and prospects. *Theoretical and Applied Genetics* 132:3227-3243, https://doi.org/10.1007/s00122-019-03433-x

Chin E. C. L., H. Shu, J. S. C. Smith and M. L. Senior (1996) Maize simple repetitive DNA sequences: abundance and allele variation. *Genome* 39:866-873, https://doi.org/10.1139/g96-109

CIMMYT, International Maize and Wheat Improvement Center (2005) Laboratory Protocols: CIMMYT Applied Molecular Genetics Laboratory. Third edition. International Maize and Wheat Improvement Center. México, D. F. 81 p.

Dotson S. B., D. A. Somers and B. G. Gegenbach (1989) Purification and characterization of lysine-sensitive aspartate kinase from maize cell cultures. *Plant Physiology* 91:1602-1608, https://doi.org/10.1104/pp.91.4.1602

Dotson S. B., D. A. Frisch, D. A. Somers and B. G. Gegenbach (1990) Lysine-insensitive aspartate kinase in two threonine overproducing mutants of maize. *Planta* 182:546-552, https://doi.org/10.1007/BF02341030

Galili G. (1995) Regulation of lysine and threonine synthesis. *The Plant Cell* 7:899-906, https://doi.org/10.1105/tpc.7.7.899

Galili G., R. Amir and A. R. Fernie (2016) The regulation of essential amino acid synthesis and accumulation in plants. *Annual Review of Plant Biology* 67:153-178, https://doi.org/10.1146/annurev-arplant-043015-112213

Gao S., C. Martínez, D. J. Skinner, A. F. Krivanek, J. H. Crouch and Y. Xu (2008) Development of a seed DNA-based genotyping system for marker-assisted selection in maize. *Molecular Breeding* 22:477-494, https://doi.org/10.1007/s11032-008-9192-4

Gupta H. S., R. Babu, P. K. Agrawal, V. Mahajan, F. Hossain and N. Thirunavukkarasu (2013) Accelerated development of quality protein maize hybrid through marker-assisted introgression of *opaque-2* allele. *Plant Breeding* 132:77-82, https://doi.org/10.1111/pbr.12009

Gutiérrez-Rojas A., J. Betrán, M. P. Scott, H. Atta and M. Menz (2010) Quantitative trait loci for endosperm modification and amino acid contents in quality protein maize. *Crop Science* 50:870-879, https://doi.org/10.2135/cropsci2008.10.0634

Hartings H., M. Maddaloni, N. Lazzaroni, N. Di Fonzo, M. Motto, F. Salamini and R. Thompson (1989) The *O2* gene which regulates zein deposition in maize endosperm encodes a protein with structural homologies to transcriptional activators. *The EMBO Journal* 8:2795-2801, https://doi.org/10.1002/j.1460-2075.1989.tb08425.x

Jarret H. W., K. D. Coosky, B. Ellis and J. M. Anderson (1986) The separation of *o*-phthalaldehyde derivatives of amino acids by reversed-phase chromatography on octylsilica columns. *Analytical Biochemistry* 153:189-198, https://doi.org/10.1016/0003-2697(86)90079-5

Laemmli U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227:680-685, https://doi.org/10.1038/227680a0

Mariotti F. and C. D. Gardner (2019) Dietary protein and amino acids in vegetarian diets —A review. *Nutrients* 11:2661, https://doi.org/10.3390/nu11112661

Maqbool M. A., A. Beshir and E. S. Khokhar (2020) Doubled haploids in maize, deployement, and challenges. *Crop Science* 60:2815-2840, https://doi.org/10.1002/csc2.20261

Mertz E. T., L. S. Bates and O. E. Nelson (1964) Mutant gene that changes protein composition and increases lysine content of maize endosperm. *Science* 145:279-280, https://doi.org/10.1126/science.145.3629.279

Nuss E. T. and S. A. Tanumihardjo (2011) Quality protein maize for Africa: closing the protein inadequacy gap in vulnerable populations. *Advances in Nutrition* 2:217-224, https://doi.org/10.3945/an.110.000182

Orita M., H. Iwahana, H. Kanazawa, K. Hayashi and T. Sekiya (1989) Detection of polmorphism of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proceedings of the National Academy of Sciences of the United States of America* 86:2766-2770, https://doi.org/10.1073/pnas.86.8.2766

Pukalenthy B., D. Manickam, K. Adhimoolam, S. G. Mahesh, N. Ramanathan, S. Chandran, … and S. Natesan (2020) Marker aided introgression of *opaque 2* (*o2*) allele improving lysine and tryptophan in maize (*Zea mays* L.). *Physiology and Molecular Biology of Plants* 26:1925-1930, https://doi.org/10.1007/s12298-020-00857-4

Rasheed A., Y. Hao, X. Xia, A. Khan, Y. Xu, R. K. Varhney and Z. He (2017) Crop breeding chips and genotyping platforms: progress, challenges, and perspectives. *Molecular Plant* 10:1047-1064, https://doi.org/10.1016/j.molp.2017.06.008

SAS Institute (2002) SAS/STAT User's Guide. Version 8. Sixth edition. SAS Institute. Cary, North Carolina, USA. 112 p.

Schmidt R. J., F. A. Burr, M. J. Aukerman and B. Burr (1990) Maize regulatory gene opaque-2 encodes a protein with a "leucine-zipper" motif that binds to zein DNA. *Proceeding of the National Academy of Sciences of the United States of America* 87:46-50, https://doi.org/10.1073/pnas.87.1.46

Semagn K., R. Babu, S. Hearne and M. Olsen (2014) Single nucleotide polymorphism using kometitive allele specific PCR (KASP): overview of the technology and its application in crop improvement. *Molecular Breeding* 33:1-14, https://doi.org/10.1007/s11032-013-9917-x

Singleton W. R. (1939) Recent linkage studies in maize. *Genetics* 24:59-63, https://doi.org/10.1093/genetics/24.1.59

Sunnucks P., A. C. C. Wilson, L. B. Beheregaray, K. Zenger, J. French and A. C. Taylor (2000) SSCP is not so difficult: the application and utility of single-stranded conformation polymorphism in evolutionary biology and molecular ecology. *Molecular Ecology* 9:1699-1710, https://doi.org/10.1046/j.1365-294x.2000.01084.x

Thompson J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin and D. G. Higgins (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 25:4876-4882, https://doi.org/10.1093/nar/25.24.4876

Vasal S. K. (2000) The quality protein maize story. *Food and Nutrition Bulletin* 21:445-450, https://doi.org/10.1177/156482650002100420

Vasal S. K. (2001) High quality protein corn. *In*: Speciality Corns. A. R. Hallauer (ed.) CRC Press LLC. Boca Raton, Florida, USA. pp:85-129.

Vivek B., A. F. Krivanek, N. Palacios-Rojas, S. Twumasi-Afriyie and A. O. Diallo (2008) Breeding Quality Protein Maize (QPM): Protocols for Developing QPM Cultivars. CIMMYT. México, D. F. 50 p.

Wang X. and B. A. Larkins (2001) Genetic analysis of amino acid accumulation in *opaque-2* maize endosperm. *Plant Physiology* 125:1766-1777, https://doi.org/10.1104/pp.125.4.1766

Wang X., D. K. Stumpf and B. A. Larkins (2001) Aspartate kinase 2. A candidate gene of a quantitative trait locus influencing free amino acid content in maize endosperm. *Plant Physiology* 125:1778-1787, https://doi.org/10.1104/pp.125.4.1778

Wang X., J. A. Lopez-Valenzuela, B. C. Gibbon, B. Gakiere, G. Galili and B. A. Larkins (2007) Characterization of monofunctional aspartate kinase genes in maize and their relationship with free amino acid content in the endosperm. *Jounal of Experimental Botany* 58:2653-2660, https://doi.org/10.1093/jxb/erm100

Xu Y. and J. H. Crouch (2008) Marker-assisted selection in plant breeding: from publications to practice. *Crop Science* 48:391-407, https://doi.org/10.2135/cropsci2007.04.0191

Zhang S., J. Zhan and R. Yadegari (2018) Maize *opaque* mutants are no longer so opaque. *Plant Reproduction* 31:319-326, https://doi.org/10.1007/s00497-018-0344-3